EMPLOYING BIG DATA TOOLS TO INQUIRE INTO ALBANIA'S RESIDENTIAL HOUSING PRICE TRENDS

An Application of Python and MS Excel's Power toolset

Orion Garo
Research Department
Bank of Albania

Prepared for the "16th South-Eastern European Economic Research Workshop",
Bank of Albania,
5 – 6 December, 2022.

RESEARCH FOCUS

Big Data:

- Use of web-residing User-Generated Content to obtain quantifiable microdata about residential housing prices in Albania.

Motivation behind this research project:

- Trends of Albania's residential housing market: difficult to track/estimate;
- No official statistics, with regard to per-square-meter prices of any real estate properties, supplied.

Objective:

- Estimating a monthly index of the housing properties' per-square-meter price for the case of Albania.
  - Intended to be achieved by:
    a) Making use of website-listed prices for housing property sales in Albania's capital, as well as its coastal and inland areas;
    b) Employing Big Data information extraction and organization tools – i.e: Python and MS Excel's Power toolset – in order to identify, gather, clean, organize, and calculate this website-residing data, so as to synthetize the price index we are interested in.

DATA ASSESSMENT

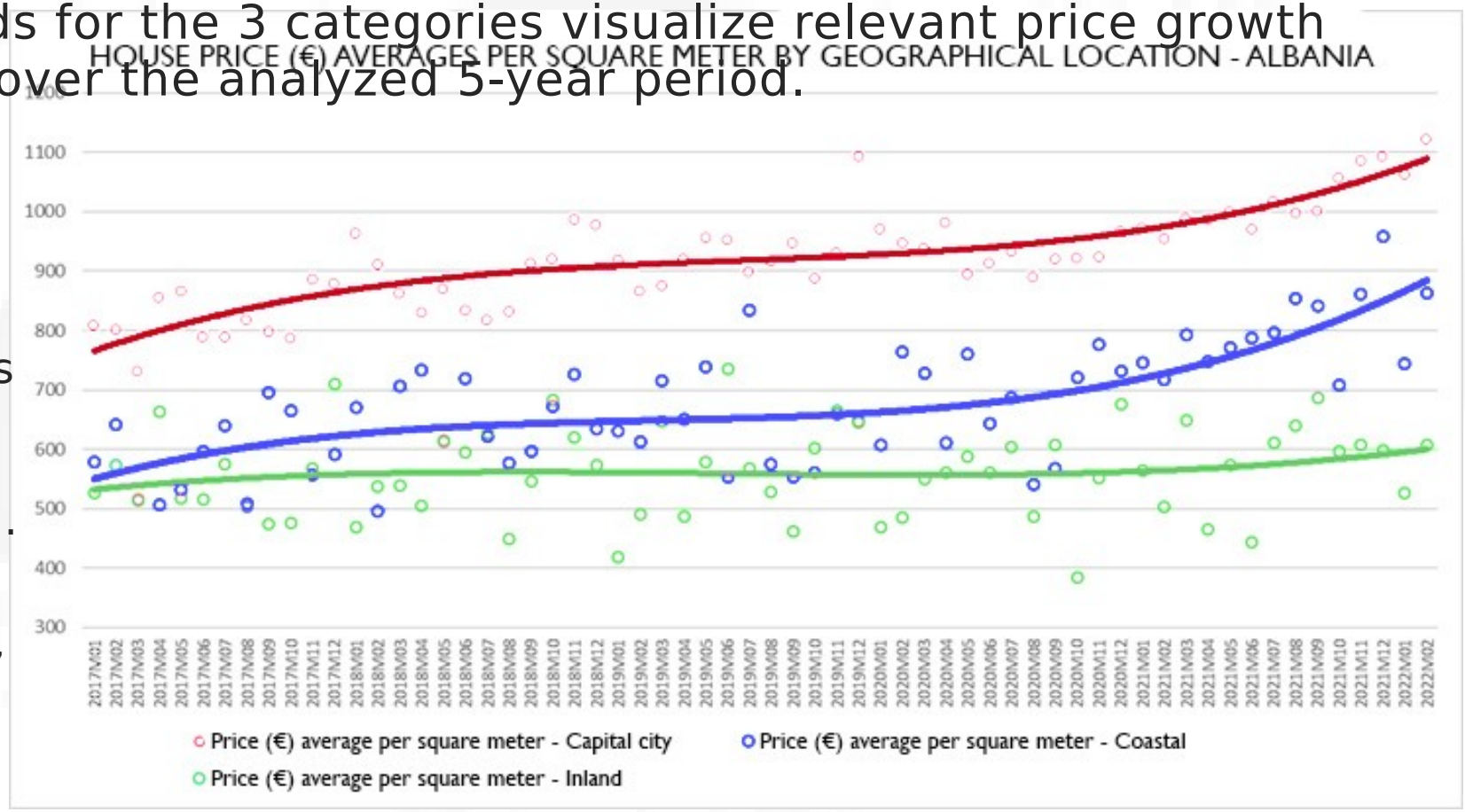
Raw data source:	njoftime.com	*Albania's largest UGC transactions information website.*
Smaple size:	29.000 records	
Sales prices in Tirana:	87.7%	
Sales prices in coastal areas:	8.3%	
Sales prices in inland areas:	3.9%	
Observation inclusion:	apartments	*in urban and semiurban apt. bulidings.*
	single storey houses/villas multi storey houses/villas	*in rural and semiurban areas.*
Time frequency:	monthly	*depending on observations' abundance, could be quarterly, weekly, daily, etc.*
No. of periods:	65	
Observations per period (month):	446	*(approximately).*
Categorizing of available data:	capital city	*sales web ads in: Tirana;*
	coastal	*Albania's coastal cities, towns and rural areas;*
	inland	*Albania's inland cities, towns and rural areas.*

METHODOLOGICAL FRAMEWORK

1. Tools used for the electronic data gathering procedure: [i] Python framework; [ii] Scrapy (a sub-component of the framework);
   - Raw data is assembled & kept in a .json file;
2. Tools used for cleaning, organizing, tabular ordering, outlier and missing values removing: [i] formulas of the Excel framework; [ii] Power Query and [iii] Power Pivot (sub-components of the framework);
   - Filtered sets, visual reports and charts, for the selected time frequency are obtained;
3. Tools used for microdata regression fitting: Eviews;
   - Tabulated data variables are fitted into an OLS regression to obtain date dummy coefficients and estimate the index.

DATA VISUAL REPRESENTATION

- Data points in this graph: calculated monthly averages of residential housing listed prices.
- Fitted trends for the 3 categories visualize relevant price growth dynamics, over the analyzed 5-year period.
- Comparable growth path trending upwards, for Tirana and coastal areas (a 200-250 Euro discrepancy).
- Prices for inland areas, markedly inert.



METHODOLOGICAL FRAMEWORK

- The econometric instrument used in this study is based on Kristo and Bollano's (2012) methodological groundwork on hedonic prices, and specifics of estimating the index are obtained from NasserEddine's (2017) hedonic regression approach.
- Ease of estimation asks for a semi-logarithmic form of the OLS specification, as detailed below:

$$\text{Log } p_{it} = \alpha + ++$$

where:

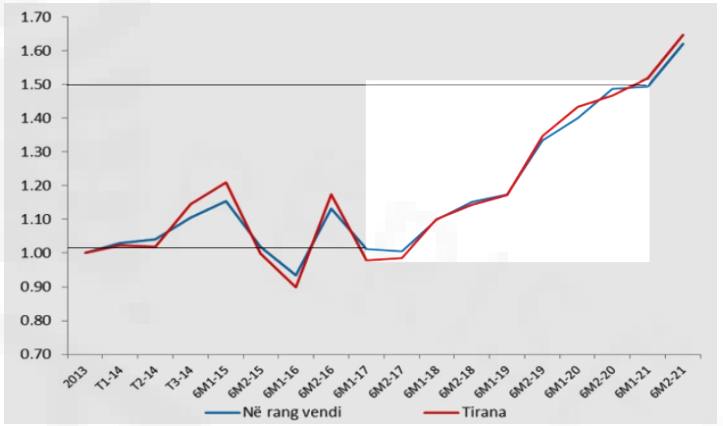
- p_{it} -> price of property i at month t;
- X_{iq} -> estimated value of the quality q of property i;
- D_{it} -> set of time dummy regressors pertaining to the listed price of property i;
- ε_{it} -> white noise residual of property i at month t.

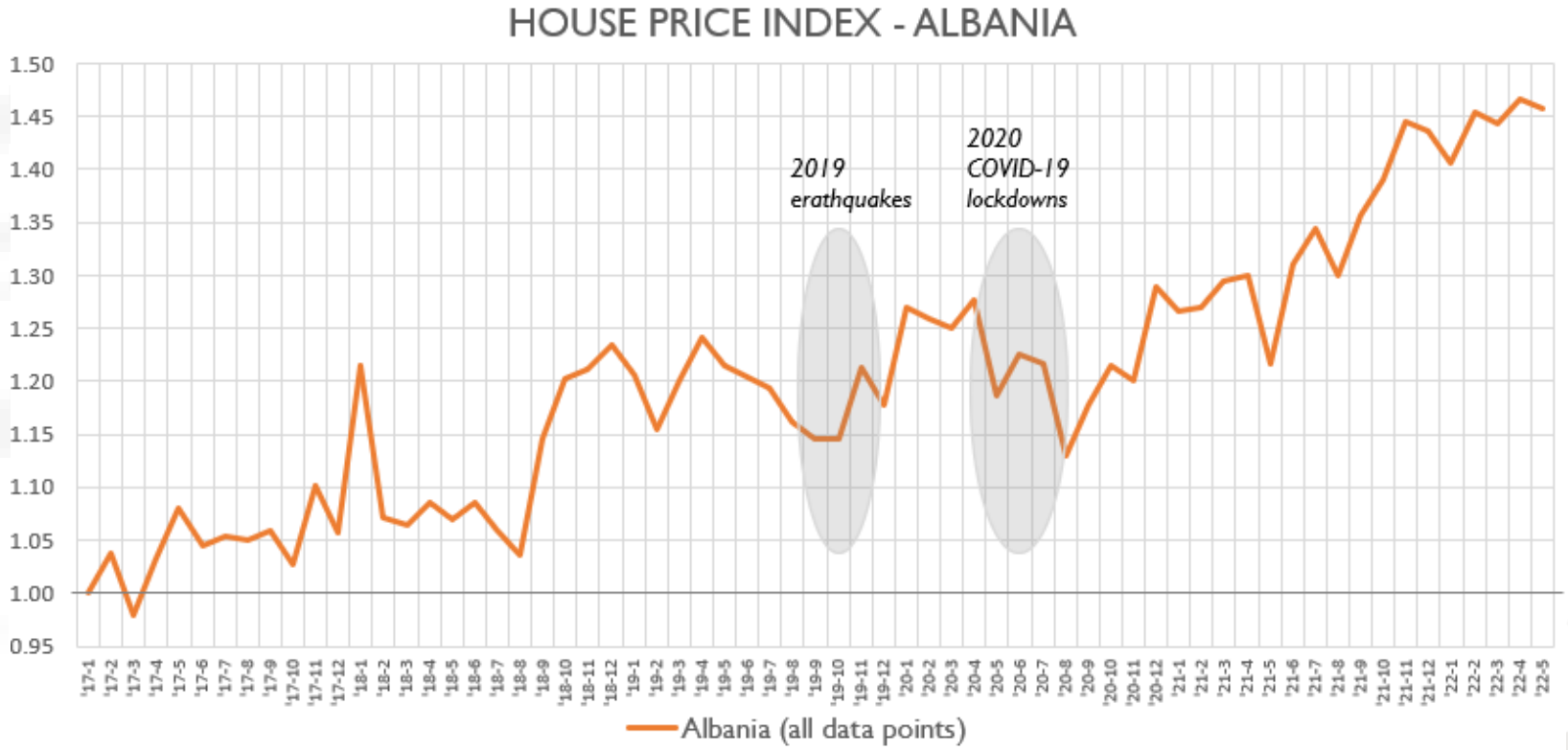
- Component X_{iq} of the estimation can allow for as many qualities of the property as raw data can provide ([i] square meters, and [ii] studio apartment or not, used for this estimation).
- Once the model is fitted and the date dummy coefficients are obtained, the index is built by keeping the initial period as the reference period, and taking the antilogs of all other periods' coefficients (NasserEddine, 2017).

OBTAINED RESULTS

- The large graph below plots the study's index.
- It displays a constant and steady increase of prices, growing about 45% over a 5 year period since beginning of 2017; capturing also the back-to-back major economic shocks of 2019 and 2020.
- As a comparison, the upper-right graph is the Fischer index of house prices for Albania.

(published by Bank of Albania's Financial Stability Department in its "Real Estate Market" questionnaire results)





OBTAINED RESULTS

- This graph decomposes the country-wide index by geographical location.
- Constant and steady increase of house prices in Tirana (about 40 percentage points over a 5-year period)
- House prices in coastal areas show relative inertia up to the first half of 2020, then jump about 30 percentage points, over 1 and a half years.
- House prices in inland areas remain inert the whole analyzed period.